5. SEGMENTATION

Segmentation is a fundamental image processing task in which pixels within each segment are homogeneous with respect to some predicate, and hence can be used to denoise the image while preserving structures. The boundaries between segments correspond directly to discontinuities in the image and hence edge detection is achievable in the same framework. It would seem obvious to apply these ideas directly to the sub-images used in local image processing. Strangely, much of the literature seems to treat local techniques differently, preferring to use adhoc methods instead, which usually require image dependent parameters to be supplied, or estimated from training sets. Segmentation algorithms which use spatial information can produce better results than others. Thresholding techniques are therefore an attractive option for local segmentation. Other image segmentation algorithms are attempted to partition an image into separate groups of related pixels, called segments. By considering the pixels similar values are related and segmented accordingly. Segmentation is also carried out in form of clustering which ignores spatial information. Although segmentation of arbitrary images is inherently difficult, the principles described have been applied successfully to Indian Sign Language images.

5.1 Approaches to Quantitative Segmentation Evaluation

The vast majority of the quantitative approaches are basically empirical discrepancy methods, analyzing the number of misclassified pixels in relation to reference segmentations. In contrast, other algorithms directly address over and under segmentation by considering the number of segments. An intersection image of the segmentation result and morphologically dilated binary reference segmentation are used to quantify the number of misclassified pixels. Other evaluation approaches are designed to minimize or exclude the a priori knowledge and the subjective (human) bias is added to the evaluation by manually created references. An approach that comprises both analytical and empirical criteria is made possible by defining a multidimensional fitness-cost-space instead of a single discrepancy-parameter-space.
A promising co-evaluation framework which combines the results of various evaluation approaches and its potentiality are explained in this chapter.

The sheer variety of algorithms are examined in terms of the following attributes:

- whether or not spatial information is used
- suitability to small sub-images
- ability to detect the number of segments automatically
- underlying segment and noise model
- region growing (identify homogeneity) or edge following (identify heterogeneity)
- attempt to optimize an objective criterion
- time and space complexity
- Parallel or sequential execution.

### 5.2 Comprehensive analysis of segmentation methods for ISL recognition

There are many unsupervised and supervised segmentation algorithms, which use low-level features, e.g. intensity and texture, to generate homogeneous patches from an input image. Four categories of segmentation are: histogram based methods, for example, the peaks, valleys and curvatures of the smoothed histogram, clustering based methods, where the gray level samples are clustered in two parts as background and foreground (object), entropy based methods that use the entropy of the foreground and background regions, object attribute-based methods which searches a measure of similarity between the gray level and the binarized images. The research is carried out to quantify the segmentation methods based on the similarities and discontinuities of the pixels including the intensity and the object structures.

#### 5.2.1 Extracting foreground from background for ISL images

Two basic types of segmentation exist at present -realized with respect to the intensity and to the intensity gradient, and two basic types of segments - areas and borders, respectively. The term “area” usually denotes topologically joined regions of the image which have comparatively homogeneous distribution of intensity, while the term “border” relates to zones where the intensity changes sharply, or in other words,
zones with greater value of the intensity gradient. Borders may either be situated between an object and the background or between different regions of the object. One of these two types of segmentation is usually applied for the purposes of image processing - the intensity or gradient which, finally leads to partial use of the intensity characteristics of the picture. That is why adaptive threshold-gradient method is compared with histogram method in this session which treats the image as one indivisible structure containing areas and borders. The analysis of this structure can be given as the result of the segmented image.

Figure 5.1: Segmentation Analysis for threshold based methods

Figure 5.1 shows the analysis of threshold based methods. Images often are of poor quality due to storage conditions, and based on the quality of capturing mode. As a result, it is difficult to separate the foreground and background. The problem is particularly difficult because many images have varying contrast, smudges, variable background intensity etc. The binarization method converts each object in the image from a matrix of pixels to a smaller number of features which therefore determines more efficiently. Every preliminary test is always done between the two main groups for thresholding, namely:

- global (histogram-based)
- adaptive (binary based)
5.2.1.1 Binary based Methods

A global method tends to binarize the image with a single threshold. Among most powerful global techniques, Otsu’s algorithm can achieve high performance with simple backgrounds and without tuning of parameters. Thresholding identifies and extract an object on the basis of distribution of gray levels or textures.

5.2.1.1.1 Iterative thresholding

The threshold value is dynamically determined for a particular segment of an image using distribution of pixel intensity. It thresholds the intensity image by returning an indexed image. The threshold values are set between 0 and 1.

Figure 5.3: Binary threshold methods (a) Original image (b) Iterative method (c) Local Threshold method (d) Otsu’s method

Figure 5.2 depicts the taxonomy and figure 5.3 gives the subjective evaluation of binary threshold methods.
5.2.1.2 Local threshold method

A single threshold value is assigned to every image and converts the grayscale image into a binary one. The output image replaces all pixels in the input image with greater luminance level with the value 1 (white) and replaces all other pixels with the value 0 (black). Level in the range [0,1] is specified in which the argument is computed automatically regardless to class of the input image.

5.2.1.3 Global image threshold using Otsu's method

A global threshold is computed to convert an intensity image to a binary image and that is a normalized intensity value level that lies in the range [0, 1]. In Otsu's method, the threshold is chosen to minimize the intraclass variance of the black and white pixels. The lower bound is attainable only by images having a single gray level, and the upper bound is attainable only by two-valued images.

5.2.1.2 Histogram based method

Often elementary and heuristic methods are used to improve images in some sense. The simplest method, called threshold method, consists of associating each element in the scene with one of the two groups: the group of the object or background depending on whether the intensity of the element exceeds a given threshold value or not (Ping Sung Liao et al., 2001). The main problem in using this method is choosing the threshold of separation. A widespread approach to the problem defines the selection of a threshold value, corresponding to the local minimum of the intensity histogram, which has to be bimodal. The chief disadvantages of the threshold methods are obtaining false regions and the loss of regions as well. When using gradient based segmentation, it is assumed that the shape of the object is fixed by its borders. The main limitation in the gradient methods is the appearance of false contours, their splitting and loss. Their advantage is avoiding the low-frequency noise in the image, i.e., the uneven illuminance. Human vision is sensitive to the contrast between the separate intensity areas and automatically ignores the irregular illuminance. The gradient methods take into account this feature of human vision and hence they are better than the threshold ones.
5.2.1.2.1 Contrast Limited Adaptive Histogram Equalization

CLAHE method is used for segmentation based on the contrast of the grayscale image and it is done by transforming the values using contrast-limited adaptive histogram equalization (Antonis Daskalakis et al., 2007). It operates on small regions in the image, called tiles, rather than the entire image. Each tile's contrast is enhanced, so that the histogram of the output region approximately matches the histogram specified by the 'Distribution' parameter. The neighboring tiles are then combined using bilinear interpolation to eliminate artificially induced boundaries. The contrast, especially in homogeneous areas is limited to avoid amplifying noise that might be present in the image. It is a special case of the histogram equalization technique that functions adaptively on the image to be enhanced. Application of CLAHE individually maximizes the contrast throughout the spots by adaptively enhancing the contrast of each cell-pixel relative to its local neighbourhood. The procedure for enhancing individual cell-images by employing the CLAHE technique is described below:

Step 1: Each cell-image has been divided into a number of non-overlapping contextual regions of equal sizes, experimentally set to be 2x2, which corresponds to approximately 40 pixels.

Step 2: The histogram of each contextual region has been calculated.

Step 3: A clip limit, for clipping histograms, has been set (t=0.001). The clip limit should be a threshold parameter by which the contrast of the cell-image could be effectively altered; a higher clip limit increases cell-spot contrast.

Step 4: Each histogram has been redistributed in such a way that its height does not exceed the clip limit.

Step 5: All histograms be modified by the transformation function and represented as eqn (5.1)

\[ T(r_k) = \sum_{j=0}^{k} P_r(\tau_j) \]  

(5.1)

Where eqn (5.2) states

\[ P_r(\tau_j) = \frac{n_j}{n} \]  

(5.2)
is the probability density function of the input image grayscale value \( j \), \( n \) is the total number of pixels in input image and \( n_j \) is the input pixel number of grayscale value \( j \).

Step 6: The neighboring tiles have been combined using bilinear interpolation and the cell-image grayscale values are altered according to the modified histograms.

![Figure 5.4: Comparison of threshold methods](image)

**Figure 5.4: Comparison of threshold methods** (a) Original image (b) CLAHE method (c) Otsu’s method

Figure 5.4 gives the comparison results of threshold methods when applied to ISL images. Traditional segmentation is computationally intensive for individual pixels. A new segmentation method is proposed which takes advantage of thresholding to increase the feature extraction rate with reasonable range as the segmented part is refined with best global method. The aim is achieved by comparing the segmentation methods with minor differences that articulates to the characteristics of an image. It concludes better segmentation method with less threshold values. The application of the method proposed is for analysis of two dimensional images with arbitrary location of the illuminating source, for reducing the number of the intensity levels and removing the information redundancy.

### 5.2.2 Image Segmentation using Contour Methods

The most popular approach for sign detection is active contour models or snakes (Hongzhi Wang. and John Oliensis, 2010). This model evolves a curve that subjects to constraints from a given image \( \mu_0 \). It contains an edge detector which depends on the gradient of the image that carries a disadvantage that is difficult to
associate with topological transformations (Giuseppe Papari. and Nicolai Petkov. 2011). Contours can be distinguished from edges. Edges are variations in intensity in a gray level image whereas contours are salient coarse edges that belong to objects and region boundaries in the image (Jagadish, H. Pujar. et. al., 2010). The contour map drawn by human observers include these edges as they are considered to be salient. However, the contours produced by different humans for a given image are not identical when the images are of complex such as natural scenes (ChengEn Lu et al., 2010). In such images, multiple cues are available for the human visual system (HVS) - low level cues such as coherence of brightness, texture or continuity of edges, intermediate level cues such as symmetry and convexity, as well as high level cues based on recognition of familiar objects. Even if two observers have exactly the same set of cues, they may choose contours at varying levels of granularity. Thus saliency of an edge is a subjective matter and varies accordingly. In general, a contour map is an efficient representation of an image since it retains only salient information and hence is more valuable for high level computer vision tasks (Dongjiang Xu and Takis Kasparis, 2003). The design of a detector that can extract all contours from a wide range of images is therefore of interest. The key to extracting contours appears, from the ground truth, to be the ability to assess what is relevant and what is not in a local neighbourhood. An assessment-based strategy has been attempted to contour detection using local information around an edge such as image statistics, topology, texture, colors, edge continuity and density. Figure 5.5 represents the taxonomy of contour based segmentation.

![Taxonomy of Contour detection](image)

Figure 5.5: Taxonomy of Contour detection
5.2.2.1 Classical active contour method in Image Segmentation

Conventional snake is a curve \( y(s) \in \mathbb{R}^2, s \in [0,1] \), which moves over the image domain to minimize its energy function (Michael Randolph Maire, 2009). The first and second order derivative regularization force the snake to have the smallest length and the smoothest shape. The contour is represented as a mean curve of the probability density function \( p(x) \) as given in eqn (5.3):

\[
p(x) = \frac{1}{2\pi\sigma^2} \exp \left( -\frac{\|x-y(z)\|^2}{2\sigma^2} \right) ds \tag{5.3}
\]

Applying this method, it is numerically necessary to keep the evolving level set function close to a signed distance function. Reinitialization of a technique for periodically re-initializing the level set function to a signed distance function during the evolution, is used for maintaining stable curve evolution. The computational bottleneck of the classical method is in solving the linear system of equations and that is achieved by multiphase function.

5.2.2.2 The Multiphase Level Set Method in Image Segmentation

A level set method is a numerical and theoretical tool for propagating interfaces (Pablo Arbelaez, et al., 2011). In image processing, it is used for propagating curves in 2D or surfaces in 3D. In this image segmentation phase, the level set method has some advantages over to the active contour model. The level set method conquers the difficulties of topological transformations and handles the complex topological changes automatically. This method takes the advantage of chanvese model in which it detects the contours both with and without gradients (Tony, F. Chan. and Luminita, A. Vese, 2001). In addition, by using this model and its level set formulation, interior contours are automatically detected even when the initial curve is present anywhere in the image (Lin He and Stanley Osher, 2007). This framework is applied to see the way in which the different image properties can be used for segmentation. In order to segment images into more regions, a Multiphase Level Set Method has also been developed. Based on the Four-Color Theorem, only four colors are adequate to dye all
the regions in a partition. Therefore only two level set functions will suffice to represent any partition. The formulation used for this method which will minimize the energy function is defined as follows in eqn (5.4)

\[
F(c, \phi) = \int_{\Omega} (u_0 - c_{11})^2 H(\phi_1) H(\phi_2) dxdy
+ \int_{\Omega} (u_0 - c_{10})^2 H(\phi_1)(1 - H(\phi_2)) dxdy
+ \int_{\Omega} (u_0 - c_{01})^2 (1 - H(\phi_1)) H(\phi_2) dxdy
+ \int_{\Omega} (u_0 - c_{00})^2 (1 - H(\phi_1))(1 - H(\phi_2)) dxdy
+ \mu \int_{\Omega} |\nabla H(\phi_1)| + \mu \int_{\Omega} |\nabla H(\phi_2)|,
\]

Where eqn (5.5) is

\[
c_{11}(\phi) = \text{mean}(u_0) \text{ in } \{(x, y) : \phi_1(t, x, y) > 0, \phi_2(t, x, y) > 0\},
c_{10}(\phi) = \text{mean}(u_0) \text{ in } \{(x, y) : \phi_1(t, x, y) > 0, \phi_2(t, x, y) < 0\},
c_{01}(\phi) = \text{mean}(u_0) \text{ in } \{(x, y) : \phi_1(t, x, y) < 0, \phi_2(t, x, y) > 0\},
c_{00}(\phi) = \text{mean}(u_0) \text{ in } \{(x, y) : \phi_1(t, x, y) < 0, \phi_2(t, x, y) < 0\}.
\]

Figure 5.6: Evaluation of Segmentation results (a) Original image (b) Local Active method (c) Global active method

Figure 5.6 gives the evaluation results of contour based segmentation methods. Active contour is adopted as the segmentation technique that explores efficient feature extraction. The drawbacks of the classical approach are the time and space complexities (Antonio Marin-Hernandez and Michel Devy). This phenomenon may cause some error in calculating boundary area. To overcome these drawbacks, multiphase level set using Chan-Vese model is used as it can deal with images with or without edges. Experiments are carried out on the methods with various sign language
images. The results show that multiphase level set is better for feature extraction and is proved both objectively and subjectively.

5.2.3 The Level set method in image segmentation

The level set method is a numerical and theoretical tool for transmitting interfaces (Yue, Y. and Tagare, H. D, 2008). It is basically used to track moving fronts by considering the front as the zero level set of an embedded function, called the level set function. (Dietenbeck,T et al., 2010). In this section, experimental study is done on level set local minima by comparing them with two datasets especially for sign language recognition. To the knowledge, such a comparison has not yet appeared in the literature especially for ISL dataset. There are two ways in which minima seeking algorithms may be compared: the value of the energy function at the minima can be compared, and the location of the minima can be compared (Nassir, H. Salman, 2009). The former is taken into account that explains the relative power of the algorithms to minimize the energy function. Figure 5.7 depicts the model for contour based hand segmentation.

![Contour based Hand Segmentation](image)

**Figure 5.7: Contour based Hand Segmentation**

5.2.3.1 Caselles method

This approach is used to connect the classical “snakes” based on energy minimization and geometric active contours with the theory of curve evolution. This algorithm is a contour-based method i.e. the gradient of the image is used to compute the force function. The curve is thus being driven to regions with high gradient. This method does not require any regularization term as it is intrinsic to the method, and eqn (5.6) represents the same
\[ \frac{\partial \phi}{\partial t}(x) = g(I(x)) \nabla \phi(x)(c + k) + \nabla_s(I(x)) \nabla \Phi(x) \quad (5.6) \]

Where \( k = \text{div} \left( \frac{\nabla \phi(x)}{\| \nabla \phi(x) \|} \right) \) corresponds to the curvature of the evolving contour and \( c \) is a constant that acts as a balloon force. Energy criterion can be defined as eqn (5.7) and \( g(I) \) is represented as eqn (5.8)

\[
E(T) = \int_0^1 g(I(\Gamma(q))) H^-(q) H dq 
\]

\[
g(I) = \frac{1}{1 + \| \nabla (G * I) \|^2} 
\]

Where \( I \) corresponds to the image intensity, \( \Gamma \) is the parametric curve.

### 5.2.3.2 Chan-Vese method

This method is applied by taking the advantage of detaching the contours both with and without gradients (Robert Crandall, 2009). In addition, by using this model and its level set formulation, interior contours are automatically detected. The formulation used for this method which will minimize the energy function is defined as follows in eqn (5.9)

\[
F(c, \phi) = \int_\alpha (u_0 - c_0)^2 H(\phi) H(\phi_1) dx dy + \int_\alpha (u_0 - c_0)^2 H(\phi_1) (1 - H(\phi)) dx dy \\
+ \int_\alpha (u_0 - c_0)^2 (1 - H(\phi)) H(\phi_1) dx dy \\
+ \int_\alpha (u_0 - c_0)^2 (1 - H(\phi_1)) (1 - H(\phi)) dx dy \\
+ \mu \int_\alpha [\nabla H(\phi_1)] + \mu \int_\alpha [\nabla H(\phi)] 
\]

### 5.2.3.3 Cumming Li method

This method is used with an energy criterion that is defined as eqn (5.10)

\[
E(\phi) = \lambda_1 \int_\alpha \int_\alpha K_\sigma(x - y) |I(y) - f_1(x)|^2 H(\phi(x)) dy dx \\
+ \lambda_2 \int_\alpha \int_\alpha K_\sigma(x - y) |I(y) - f_2(x)|^2 (1 - H(\phi(x))) dy dx \\
+ \nu \int_\alpha |I(\phi(x))| \| \nabla \phi(x) \| dx + \mu \int_\alpha \frac{1}{2} (\| \nabla \phi(x) \| - 1)^2 dx 
\]

Where \( I(x) \) is the image intensity at pixel \( x \), \( H \) is the Heaviside function, \( K_\sigma \) is a Gaussian Kernel defined as eqn (5.11):

\[
K_\sigma(u) = \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-|H - 1/2|^2} 
\]

89
With a scale parameter $\sigma > 0$, $f_1$ and $f_2$ are two functions centered at pixel $x$ and computed at each iteration as given in eqn (5.12) and (5.13):

$$f_1(x) = \frac{K_\sigma * (H(\phi(x))I(x))}{K_\sigma * H(\phi(x))}$$  \hspace{1cm} (5.12)

$$f_2(x) = \frac{K_\sigma * (1-H(\phi(x)))I(x))}{K_\sigma * (1-H(\phi(x)))}$$ \hspace{1cm} (5.13)

The first two integrals are correspond to data attached term localized around each point $x[7]$, the third to the usual regularization term that smoothes the curve during its evolution and last is a regularization term that forces the level-set to keep signed distance properties over the evolution process respectively.

5.2.3.4 Lankton method

The evolution equation used in this method is represented in eqn (5.14)

$$E\phi = \int_{\Omega} \delta(\phi(x)) \int_{\Omega} B(x, y) F(I(x), I(y) dy dx + \lambda \int_{\Omega} \delta(\phi(x)) \|\nabla \phi(x)\| \, dx$$ \hspace{1cm} (5.14)

Where $\delta$ is the dirac function, $B$ is radius, $r$ centred at point $x$ and represented in eqn (5.15):

$$B(x, y) = 1, \|x - y\| \leq r$$

0, otherwise \hspace{1cm} (5.15)

This algorithm is a region-based method and feature term is computed locally (Samuel Dambreville et al., 2008). This allows the algorithm to segment non homogeneous objects, in which the method is sensitive to initialization.

5.2.3.5 Bernard method

Let $\Omega$ be a bounded open subset of $\mathbb{R}^d$ and let $\Omega \subset \mathbb{R}^d$ be a given d-dimensional image (Bernhard Petersch, 1999). In the B-Spline level-set formalism, the evolving interface, $\Gamma C \mathbb{R}^d$ is represented as the zero level-set of an implicit function $\phi(.)$ expressed as a linear combination of B-Spline basis functions represented as eqn (5.16)

$$\phi(x) = \sum_{k \in \mathbb{Z}^d} c[k] \beta^n \left( \frac{x}{h} - k \right)$$ \hspace{1cm} (5.16)

In eqn (5.16) $\beta^n(\cdot)$ is the uniform symmetric $d$-dimensional $B$-spline of degree $n$. 
The energy criterion is expressed as eqn (5.17)

\[ E(\phi) = \int_{\Omega} F(I(x), \phi(x))dx \]  

(5.17)

and eqn (5.18) represents \( F \) as

\[ F(I(x), \phi(x)) = H(\phi(x))(I(x) - \mu)^2 + (1 - H(\phi(x)))(I(x) - \mu)^2 \]  

(5.18)

5.2.3.6 Shi method

The method implemented is a fast algorithm based on the approximation of the level-set based curve evolution (Michael Maire et al., 2008). The implicit function is approximated by a piece-wise constant function taking only four values (-3, -1, 1, 3) corresponding respectively to the interior points, the interior point adjacent to the evolving curve, the exterior points adjacent to the evolving curve and the exterior points. The two narrow-bands that enclose the evolving contours are gathered into two lists that are updated at each iteration from simple rules, making the algorithm particularly fast. The evolution equation used is the same as ChanVese and it is given by eqn (5.19)

\[ F(I(x), \phi(x)) = H(\phi(x))(I(x) - \mu)^2 + (1 - H(\phi(x)))(I(x) - \mu)^2 \]  

(5.19)

Where, \( H \) is the Heaviside function

Figure 5.8: Implementation with sample SL database and its subjective evaluation
Performance of level set methods is experimented with two datasets of sign language which is shown in figure 5.8 and 5.9. It is stated that the shi level set method works better than the compared segmentation methods based on level sets on the sign language representative images. Experimental results with six images show that the level set evolution strategy gives minima whose energy is close to the minimum energy obtained from simulated annealing and random search. (Yong Yue Hemant D. Tagare).

5.2.4. Topological derivative based image segmentation

The visual relevance of the segmentation problems need to be considered and that issue is effectively handled based on topological techniques (David Letscher and Jason Fritts, 2007). Figure 5.10 gives the taxonomy for segmentation based on topological derivative method.
Topological derivative quantifies the sensitivity of the above problem when the domain is perturbed by the introduction of heterogeneity (hole, inclusion, source, term) (Machado, D.A. et al., 2010). Figure 5.11 depicts the concepts of topological method. It is considered $\Omega$ to be a bounded open set in $\mathbb{R}^N$ ($N=2,3$) and $\gamma_\varepsilon$ be a crack of the $\varepsilon$ centered at point $x_0$.

$$\psi_\varepsilon(\Omega_\varepsilon) = \psi(\Omega) + f(\varepsilon)D_\tau(x) + 0f((\varepsilon))$$ (5.20)

where $f(\varepsilon)$ is known positive function which going to be zero with $\varepsilon$, and $D_\tau(x)$ is the topological derivative at point $x$ given in eqn (5.20).

### 5.2.4.1 Discrete topological derivative

This algorithm used is based on discrete topological derivative concept proposed by larrabide in which the cost functional used for discrete approach is represented by the eqn (5.21) (Larrabide. et al., 2005)

$$\Psi(\Omega_\varepsilon^{\pi}) = \sum_s \sum_{pen} k^{s,p} \Delta \Omega_\varepsilon^{s,p} . \Delta \Omega_\varepsilon^{s,p}$$ (5.21)

It is given by the difference between perturbed cost function and original cost function.
5.2.4.2 Continuum topological derivative

It is same as discrete but the design vector is \( b = \{ \phi_1, \phi_2, \ldots, \phi_n \} \). This structure is modeled/analyzed as a continuum. Analytical models can therefore be large and expensive (Ma Zin Myint Maw and Khaing Khaing Aye, 2008).

![Figure 5.12: Image segmentation using topological derivative algorithm (a) Original image (b) continuum derivative (c) Discrete derivative](image)

One of the main purposes of the proposed method is to precisely segment the image without misplacing of imperative information. The filters used internally are isotropic and anisotropic filter. The approach is to prove the significance of topological derivative segmentation. The evaluation of topological segmentation methods is given in figure 5.12. The evaluation is categorized as objective and subjective in which the result over the segmentation has given hopeful outcome. The results indicate that the approach is more robust and accurate than conventional segmentation methods mainly for foreground/background segmentation evaluation problem. The computational time and number of iteration can be comparatively condensed by using optimization technique in further research.

5.2.5. Boundary based image segmentation

One of the most popular ways of localizing the hand in the image is the use of threshold process method for boundary detection and figure 5.13 shows the usage of boundary detection method for segmenting the object.
This system adopts peak detection and valley extraction algorithm to threshold the values used for histogram. The algorithm locates the globally significant peaks of the histogram (Eitan Sharon et. al., 2001). After the peaks are selected, the minimum values between any adjacent peaks are considered as valleys. The valleys are the boundaries for the segmentation. Figure 5.14 shows the implementation of boundary based method for hand detection.

The threshold is created by eqn (5.22) as

\[ B(i,j) = 1 \text{ if } F(i,j) < t \]
\[ B(i,j) = 0 \text{ if } F(i,j) \geq t \]  

(5.22)

This is a probabilistic method that makes parametric assumptions about object and background intensity distributions and then derives “optimal” thresholds. The computational demands of the method make it suitable only for filling up the gaps in a partially completed boundary. The strategy is to consider each control point in turn and move it to the pixel; in its local neighbourhood which gives the minimum. For a closed boundary it could make the initial estimate surrounding the object of interest, and add in another term to the objective function to penalize the total length. A difficulty with this type of strategy is that the control points.
Figure 5.15: Boundary based segmentation (a) Original Image (b) Labeled Image (c) Segmented Image

Figure 5.15 shows the subjective performance of the segmentation method. This method traces the exterior boundaries of objects, as well as boundaries of holes inside these objects, in the binary image. The two important needs for this method is, first, it keeps the overall complexity of the segmentation process low, and secondly, it eliminates candidates that may produce smooth continuations, but otherwise are inconsistent with the segments in the image. This simplifies the decisions made by the segmentation process and generally leads to more accurate segmentation. Hence the technique adopted is applicable which gives better input to boundary based segmentation process and also the system is made background independent.

5.2.6. Pixelwise based image Segmentation

When the segmented regions of one image are essentially indiscernible, image segmentations should be avoided to save much computation overheads. Some images confer only one semantic object and should deal with a whole rather than those segmented meaningless regions (Phung SL et al., 2005).

![Diagram of pixelwise method]

**Figure 5.16: Taxonomy of pixelwise method**
Thus, this puzzle also presents a challenge to detect those images that refuse to segmentation. So the necessity of segmentation should be evaluated according to image complexity together with the implicit region sizes and figure 5.16 shows the taxonomy of pixelwise method.

5.2.6.1 Colour based segmentation

In the past, different colour spaces have been used in skin segmentation (Nariman Habili et al., 2004). In some cases, colour classification is done using only pixel chrominance because it is expected that skin segmentation may become more robust to lighting variations if pixel luminance is discarded (Peter Gejgus et al., 2004). Investigation is done to choose the colour space and the use of chrominance channels that affect skin segmentation. Four representative color spaces commonly used in the image processing field are chosen:

RGB: Colors are specified in terms of: Red (R), Green (G), and Blue (B).
HSV: Colors are specified in terms of hue (H), saturation (S), and intensity value (V) which are the three attributes that are perceived about color. The transformation between HSV and RGB is nonlinear and similar color spaces are HIS, HLS, and HCI.
YCbCr: Colors are specified in terms of luminance (the Y channel) and chrominance (Cb and Cr channels). The transformation between YCbCr and RGB is linear. Other similar color spaces include YIQ and YUV.
CIE-Lab: Designed to approximate perceptually uniform color spaces (UCSs), and it is related to the RGB color space through a highly nonlinear transformation.

5.2.6.2 Gradient based segmentation

The segmentation method used is based on the gradient magnitude function in which its magnitude is computed by sobel operator and employed as the definition of homogeneity criterion (Stoyan Donchev, 2000). A significant rate of change of values in the immediate neighbourhood is reflected. The rate of change is precisely what is measured by the gradient of the voxel. Computing the gradients with a filter mask is done to enhance the boundary by analysis between the colon and its surrounding.
This method proves little advantage equivalent to colour based as it focuses only on the neighbourhood of the boundary, not on the whole large volumetric data, making it more computationally efficient. The gradient of a scalar function \( f(x_1, x_2, \ldots, x_n) \) is denoted by \( \nabla f \) or \( \vec{\nabla} \) where \( \nabla \) denotes the vector differential operator, \( \text{del} \). The notation \( \text{grad} \ (f) \) is also used for the gradient. The gradient of \( f \) is defined to be the vector filed whose components are the partial derivatives of \( f \) which is represented as eqn (5.23)

\[
\nabla f = \left( \frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right)
\]

(5.23)

The gradient is written as a row vector, but it is often taken to be column vector.

### 5.2.6.3 Threshold based segmentation

Thresholding can be applied locally within a neighbourhood of each pixel, or globally. Due to high varying defect sizes, it is impossible to find one neighborhood size that works for all. It starts with background removal and threshold-based object segmentation. The high-low threshold for each operation is set during the algorithm performance. It is a fixed threshold method called single band thresholding. It is perhaps the simplest segmentation method that assumes the objects to have pixel values generally different from the background. A binary output image containing only 0’s (background) and 1’s (object), is created by applying a threshold (eqn 5.24):

\[
V_{\text{out}}(i, j) = \begin{cases} 
0 & \text{if otherwise } V_{\text{in}}(i, j) < \text{threshold} \\
1 & \text{if otherwise}
\end{cases}
\]

(5.24)

This can be extended to as many different classes as necessary by defining thresholds \( t_1, t_2, t_3 \) such that:

\[
\begin{align*}
0 & \text{if } V_{\text{in}}(i, j) < t_1 \\
1 & \text{if } t_1 \leq V_{\text{in}}(i, j) < t_2 \\
2 & \text{if } t_2 \leq V_{\text{in}}(i, j) < t_3
\end{align*}
\]

(5.25)

The most common and reliable method of finding the thresholds is manual selection.
5.2.6.4 Cluster based Segmentation

Clustering refers to the process of grouping samples so that the samples are similar within each group which are called clusters (Jose Alfredo, et al., 2011). It is a way to separate groups of objects. K-means clustering is used that treats each object as having a location in space. It finds partitions such that objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible. In K-means clustering the number of clusters is partitioned and a distance metric is used to quantify how close two objects are to each other. This K means is fast iterative and leads to a local minimum which looks for unusual reduction in variance. This iterative algorithm has two steps (http://en.wikipedia.org/wiki/K-means_clustering):

Assignment step: Each observation to the cluster with the closest mean is assigned as shown in eqn (5.26)

\[ S_i^{(t)} = \{ X_j : \| X_j - m_i^{(t)} \| \leq \| x_j - m^{(t)} \| \} \] (5.26)

Update step: The new means to be centroid of the observations in the cluster is calculated as eqn (5.27)

\[ m_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} X_j \] (5.27)

Figure 5.17: Pixel wise method applied on the images (a) Original Image (b) colour based (c) Gradient based (d) Threshold based (e) Cluster based

The segmentation methods adopted are applied on a variety of images. Around 50 images are tested. The outputs are compared pixel-wise with the manually
segmented skin images. A few images are shown in the figure 5.17 for reference. In terms of color quantization, it is found that finer color quantization (a larger histogram size) gives better segmentation results. Observations on the performances of segmentation not only confirms superiority of colour pixelwise method, but also reveals that more sophisticated segmentation leads to better recognition rates.

5.2.7 Optimizing Threshold Segmentation

The problem of estimating the human gestures from a set of still images requires prospective segmentation method. Image segmentation is the central task of sign language recognition system. It is a significant step in image processing in which threshold segmentation is a simple and important method in grayscale image segmentation. The multilevel threshold image segmentation method based on maximum entropy and Particle Swarm Optimization (PSO) is presented in figure 5.18. The proposed algorithm takes the advantage of the characteristics of particle swarm optimization, with improved parameter and evolitional process. This algorithm not only considers the spatial information, but also considers the gray information and decreases the computing quantity. Several images are segmented in experiments and compared with other related segmentation methods in which few are discussed through subjective and objective assessment.

```
Detection of edges and contour of the gestures
Optimized Segmentation
Histogram based thresholding
Particle swam optimization
```

Figure 5.18: Taxonomy for optimized segmentation using PSO

5.2.7.1 Gray level based Segmentation

Graylevel segmentation is done by separating the pixel of the image into some non-intersecting regions (Shahana Bano et al., 2010). Therefore, it is very important to choose an optimal threshold value in grayscale segmentation (Ashok Kumar, D and Esther, J. 2011). By choosing an adequate threshold value object can be easily
extracted from grayscale image. Selection criterion of threshold value plays a vital role in threshold segmentation method. A PSO algorithm is adopted to solve the optimal segmentation threshold value by taking its quick convergence as the main advantage. In traditional 2D gray histogram, the horizontal axis denotes the gray-level value of pixel. It’s range is from 0 to L-1 where f(m,n) denotes the gray-level value of pixel which is located at (m,n). The range of m is from 1 to M. The range of n is from 1 to N. That is {1,2,..., M} and n {1,2,...,N} . g(m,n) is the average gray value of the neighborhoods of pixel which are located at (m,n). Object pixels and background pixels locate the diagonal neighbourhood that is A & B as noise points and edge pixels are far from diagonal that is C& D. According to associated criterion, optimal segmentation threshold value is obtained, such as a pair of values (s, t). The value s (associated criterion) represents the segmentation threshold value about pixels and value t (selection criterion) represents the segmentation threshold value about average pixels. An example for histogram is given in figure 5.19.

![Figure 5.19: Example of Histogram](image)

### 5.2.7.2 Analysis of different gray level based segmentation for human pose

The automatic selection of a robust, optimum threshold has remained a challenge in image segmentation. Besides being a segmentation tool on its own, thresholding is frequently used as one of the steps in many advanced segmentation methods. In gesture recognition applications, thresholding is not applied on the original images, but applied in a space generated by the segmentation method (Mingxin Zhang. et, al., 2008). Thus, the adopted gray level based segmentation methods are applied to few sign language images, to determine effective thresholds for further construction of the system.
5.2.7.3 OTSU’s method for segmentation

This method computes the global threshold (level) that is used to convert an intensity image to a binary image. Global thresholding consists of setting an intensity value (threshold) such that all voxels having intensity value below the threshold belong to one phase, the remainder belong to the other. Global thresholding is considered to be a good degree of intensity separation between the two peaks in the image. Eqn (5.28) explains the same

\[
\sigma_{\text{within}}^2(T) = \omega_n(T)\sigma_n^2 + \omega_o(T)\sigma_o^2(T)
\]

\[
\omega_n(T) = \sum_{i=0}^{L-1} p(i)
\]

[0, L-1] the range of intensity levels

\[
\omega_o(T) = \sum_{i=L}^{L-1} p(i)
\]

\[
\sigma_n^2(T) = \text{The variance of the pixels in the background (below threshold)}
\]

\[
\sigma_o^2(T) = \text{The variance of the pixels in the foreground (above threshold)}
\]

5.2.7.4 K-means algorithm for Segmentation

K-means is used as a two phase iterative algorithm to minimize the sum of point-to-centroid distances, summed over all k clusters. The first phase uses each iteration that consists of reassigning points to their nearest cluster centroid. The second phase uses points that are individually reassigned, arbitrarily chooses k data points to act as cluster centers. Until the cluster centers are unchanged following are the steps carried out, allocating each data point to cluster whose center is nearest and then replacing the cluster centers with the mean of the elements in their clusters end.

5.2.7.5 Delaunay algorithm for segmentation

In this segmentation, a triangular mesh is used to divide an image into several disjoint regions whose characteristics, such as intensity and texture, are similar (Michal Spanel. et. al., 2007). This concept of the vector based segmentation has a number of advantages: (1) Direct vector representation of image regions eliminates...
difficult process of raster data vectorization. (2) Easy manual corrections of the segmentation. (3) The presented two-dimensional case can be extended to the 3D space. (4) Effective representation of image structure reduces number of triangles. The Delaunay triangulation of a set of points generates regularly shaped triangles and is preferred over alternative triangulations for image segmentation. The Delaunay triangulation can be constructed by several methods and most common is the Incremental Method. Adaptive segmentation scheme is based on the Delaunay triangulation, while the mesh of the Delaunay triangulation is adapted to the underlying structure of the image data. During construction of the Delaunay triangulation, the image is divided into a number of non-overlapping triangles $t_i$. These triangles are not segments of the image by itself, but they belong to image regions $R_k = \{t_1, t_2, \ldots, t_n\}$. This relationship can be expressed by a region membership function. *Hard assignment* means that this function assigns exactly to one region to a given triangle. In practice, membership function of the form $m(t_i, R_k) = p(t_i | R_k)$ making a *soft assignment* of triangles overcomes the hard one and leads to better results. Soft membership function is usually a likelihood function assigning each triangle into every image region with some certainty. The value is higher as the similarity of the triangle and the region increases.

### 5.2.7.6 Particle Swarm Optimization for segmentation

Computation of optimal threshold is handled with Particle Swarm Optimization (PSO) which has its inspiration from the behavior of bird groups (Liping Zheng et al., 2009). It takes the particle as an individual and flight with certain speed in the search space. These particles have no quality and volume. Every particle has the simple rules of conduct. According to the flying experience of individuals and groups, particles can dynamically adjust the flying speed (Parag Puranik et al., 2011). Using PSO algorithm, optimal threshold is obtained for segmentation which is therefore implemented for sign language datasets. There are six important control parameters in PSO algorithm. They are population size cognitive learning rate $c1$、social learning rate $c2$、the maximum of particle flying speed $\max V$，the inertia weight factor $\omega$、and constriction factor $K$. 

103
The population size of the particles refers to the number of particles in iterative process. The specific model used is given in eqn (5.29)

\[
\begin{align*}
\nu_{id}(t+1) &= \nu_{id}(t) + c_1 r_1(P_{id}(t) - x_{id}(t)) + c_2 r_2(P_{gd}(t) - x_{id}(t)) \\
x_{id}(t+1) &= x_{id}(t) + \nu_{id}(t+1)
\end{align*}
\] (5.29)

The algorithm is used in this case because the sum of \(c_2\) and \(c_1\) should be 4 and the values of parameters \(c_2\) and \(c_1\) are 2. The values of the parameters \(r\) are 0.5. The iterative formula of speed is eqn (5.30).

\[
\nu_{id}(t+1) = \nu_{id}(t) + P_{id}(t) + P_{gd}(t) - 2x_{id}(t)
\] (5.30)

As the gray value of pixel is integer, the solution of SDAIVE maximum can take as integer planning problem. Pixel particles move along the gray value (V-direction) and the gray D-value (μ-direction) at the same time. Speed and location of particles need to iterate in two directions at the same time.

The figure 5.21 and 5.22 gives the visual assessment of the segmentation method based on the threshold values. Image segmentation techniques are widely used in similarity searches. The uniqueness of the proposed image segmentation algorithm includes a new quantization method to generate gray level based segmentation where the optimal threshold is automatically estimated based on particle swarm optimization. An approach is proposed that deals with the threshold value according to the similarity between gray levels. Such a similarity is obtained through a biological inspired computing technique, PSO. This method overcomes the local minima that affect most of the conventional segmentation methods. Furthermore, the proposed method is both fast and efficient to extract and the segmentation results are close to human perceptions.

![Figure 5.21: Analysis of different segmentation methods based of histogram level](image)

Figure 5.21: Analysis of different segmentation methods based of histogram level
Figure 5.22: Subjective assessment of the adopted segmentation methods

5.3 Summary

Segmentation is a challenging problem, and its accuracy implies more in the system construction. The concern is to develop a well-built segmentation work which helps to promote a better vision-based sign language recognition system, in order to provide access to future services like sign language translation. The method proposed is incorporated for object annotation purpose and its main objective is to segment the
hand with reliable segmentation during hand shape change and hand crossing. Extracting foreground from background is the essential element of sign language recognition system and it is carried out by OTSU method. By using this threshold limit method the image is read pixel by pixel and a binary image is obtained that has 1 for desired region and 0 for background. Next, detection of edges and contours of the gestures are proved by four methods. (i) Active contour (ii) Topological Derivative (iii) Boundary based (iv) Pixel wise. Optimization of threshold values is done using particle swarm optimization technique for detection of hand signs. Based on the performance measures, the color space and threshold based segmentation methods appropriately meets the requirements for segmentation process while the gradient, cluster and contour based methods only discloses the complexity of the segmentation process. This greatly facilitates the image localization process which is important for the purpose of recognition. Further, the adopted segmentation method helps the features capturing the configuration of the signers hand for improving the accuracy of the recognition system. While this chapter describes the employed and evaluated segmentation methods, the next chapter, feature extraction explains the task of extracting key vector inputs to the recognition process.